Journal of Biomolecular Techniques • Volume 33(2); 2022 Jul

Article Watch: July, 2022

Clive Arthur Slaughter

Association of Biomolecular Resource Facilities

Published on: Jul 19, 2022

DOI: https://doi.org/10.7171/3fc1f5fe.7f6f2cf1

ABSTRACT

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, AU-UGA Medical Partnership, 1425 Prince Avenue, Athens GA 30606. Tel; (706) 713-2216: Fax; (706) 713-2221: Email; cslaught@uga.edu or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the Association.

NUCLEIC ACID SEQUENCING

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V, Mikheenko A, Vollger M R, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt S J, Diekhans M, Logsdon G A, Alonge M, Antonarakis S E, Borchers M, Bouffard G G, Brooks S Y, Caldas G V, Chen N-C, Cheng H, Chin C-S, Chow W, Lima L G D, Dishuck P C, Durbin R, Dvorkina T, Fiddes I T, Formenti G, Fulton R S, Fungtammasan A, Garrison E, Grady P G S, Graves-Lindsay T A, Hall I M, Hansen N F, Hartley G A, Haukness M, Howe K, Hunkapiller M W, Jain C, Jain M, Jarvis E D, Kerpedjiev P, Kirsche M, Kolmogorov M, Korlach J, Kremitzki M, Li H, Maduro V V, Marschall T, Mccartney A M, Mcdaniel J, Miller D E, Mullikin J C, Myers E W, Olson N D, Paten B, Peluso P, Pevzner P A, Porubsky D, Potapova T, Rogaev E I, Rosenfeld J A, Salzberg S L, Schneider V A, Sedlazeck F J, Shafin K, Shew C J, Shumate A, Sims Y, Smit A F A, Soto D C, Sović I, Storer J M, Streets A, Sullivan B A, Thibaud-Nissen F, Torrance J, Wagner J, Walenz B P, Wenger A, Wood J M D, Xiao C, Yan S M, Young A C, Zarate S, Surti U, Mccoy R C, Dennis M Y, Alexandrov I A, Gerton J L, O'neill R J, Timp W, Zook J M, Schatz M C, Eichler E E, Miga K H, Phillippy A M. The complete sequence of a human genome. *Science* 376;2022:44-53.

The Telomere-to-Telomere consortium announces a major revision of the previously current human reference genome produced originally by the Genome Reference Consortium (GRC) in 2013 and updated in 2019 (GRCh38.p13). The new reference genome, designated T2T-CHM13v1.1, fills in assembly gaps, completes unfinished regions, and resolves previously incorrect assemblies. The new results are achieved in part by the use of single-molecule long-read sequencing, enabled by technology from Pacific Biosystems (Menlo Park, CA) and Oxford Nanopore Technologies (Lexington, MA), and notably by PacBio's "HiFi" circular consensus sequencing, which provided 20-kbp read-lengths with an error rate of 0.1%. These methods permitted resolution of highly repetitive centromeric satellite arrays and closely related segmental duplications. The Consortium solves the problem of allelic variation by choosing a cell line derived from a hydatidiform mole for sequencing. Mole pregnancies of this type arise by loss of the maternal complement of DNA and duplication of the paternal complement after fertilization. The present cell line indeed shows nearly uniform homozygosity with a 46,XX karyotype. This minimization of allelic variation allows correction of many inaccuracies, but the Y chromosome is absent from the new assembly. The Consortium plans to publish a revised Y chromosome sequence separately. The new sequences span ~8% of the human genome. Particularly notable is the

completion of sequences for the short arms of the acrocentric chromosomes, including the arrays of genes for ribosomal RNAs, numbering ~400 copies in the diploid genome. The present paper references further companion studies from the T2T Consortium that analyze the centromeric satellites, segmental duplications, transcriptional and epigenetic profiles, mobile elements, and variant calls. The new results enable analysis of the extensive polymorphic variation in the most repetitive regions of the genome going forward. The T2T Consortium is working closely with the Human Pangenome Reference Consortium to document the full diversity of human populations. The new assembly is particularly important too for its contributions to the structure of euchromatic segmental duplications. These long, highly similar sequences are associated with chromosomal rearrangements, and hence with the occurrence of neurodevelopmental disorders.

GLYCANS

O'Leary T R, Critcher M, Stephenson T N, Yang X, Hassan A A, Bartfield N M, Hawkins R, Huang M L. Chemical editing of proteoglycan architecture. ***Nature Chemical Biology*** **18;2022:634-642.**

O'Leary *et al*. describe methodology for the creation of semisynthetic proteoglycans for use in structure-function studies of the core protein and its membrane localization and for study of the structural features of the core protein that determines attachment of the glycosaminoglycan (GAG) chains. The authors express the protein in *E. coli*, which lacks glycosylation capability, and employ amber codon (TAG) reassignment to install an alkyne-functionalized amino acid, p-propargyl-tyrosine, at sites to which they desire to attach an azido-GAG. A *C*-terminal poly-histidine tag permits binding of the protein to nickel-nitriloacetic acid (NTA) to enable anchoring of the protein to the surface of live cells *via* a functionalized cholesterol, cholesterol-PEG2000-NTA, embedded into the phospholipid membrane. Proteins interacting with the proteoglycan are identified by tagging the proteoglycan with an ascorbic acid peroxidase (APEX) for proximity labeling. When supplied with H_2O_2, the APEX enzyme converts exogenously supplied biotin-phenol to the short-lived biotin-phenoxyl radical, which can covalently label proteins within a 20-nm radius. Proteins labeled in this way can then be purified with streptavidin and identified as presumed interactors with the proteoglycan. The authors use this assemblage of methods to study the syndecan family of proteoglycans. Syndecan molecules are naturally linked to 3-5 GAG chains of heparan sulfate and/or chondroitin sulfate. They interact with growth factors and other ligands on cell surfaces. The authors use the semisynthetic proteoglycan-mimics to dissect how structural features of proteoglycans affect ligand binding and mediate biological processes, exemplified by the differentiation of mouse embryonic stem cells and metastatic capability of human mammary carcinoma cells.

METABOLOMICS

Anglada-Girotto M, Handschin G, Ortmayr K, Campos A I, Gillet L, Manfredi P, Mulholland C V, Berney M, Jenal U, Picotti P, Zampieri M. Combining CRISPRi and metabolomics for functional annotation of compound libraries. ***Nature Chemical Biology*** **18;2022:482-491.**

Initially interested in developing an improved experimental platform for distinguishing antimicrobial drug candidates with new mechanisms of action from ones with conventional mechanisms, the authors here go on to demonstrate a general strategy for high-throughput analysis of the functionality of compounds in cells ranging from bacterial to human. Once a compound that inhibits a cellular characteristic such as growth rate or colony size has been identified, the most powerful strategy for investigation of its functionality is a screen that covers the maximum number of possible mechanisms. Rather than utilize a conventional transcriptomic or proteomic approach, the authors employ non-targeted metabolomics for their purpose because it is scalable up to very high dimensionality in terms of the numbers of features monitored; it can be used to track rapid changes in cell physiology, including changes affecting growth-essential genes; and the parameters monitored in metabolomics are operationally independent of growth-inhibition. For the construction of a reference map of metabolic changes associated with defined functional perturbations, the authors employ a screen based upon clustered regularly interspersed short palindromic repeat technology for interference in gene expression (CRISPRi). The authors document the metabolic fingerprints of 252 gene knock-downs in *E. coli*; then compare these fingerprints with the changes induced by 1,342 small-molecule drugs. They reproduce known mechanisms of existing antibiotics, and successfully predict how drugs with unknown mechanism are acting. They show that metabolic screening helps to understand the mechanism of drugs that don't bind proteins directly, and drugs that have have poor antibacterial activity. The methodology is validated in *Mycobacterium smegmatis* and in a lung cancer cell line as well as in *E. coli*. The approach is expected to be widely applicable for identifying new drugs with novel mechanisms of action.

MACROMOLECULAR SYNTHESIS & SYNTHETIC BIOLOGY

Fryszkowska A, An C, Alvizo O, Banerjee G, Canada K A, Cao Y, Demong D, Devine P N, Duan D, Elgart D M, Farasat I, Gauthier D R, Guidry E N, Jia X, Kong J, Kruse N, Lexa K W, Makarov A A, Mann B F, Milczek E M, Mitchell V, Nazor J, Neri C, Orr R K, Orth P, Phillips E M, Riggins J N, Schafer W A, Silverman S M, Strulson C A, Subramanian N, Voladri R, Yang H, Yang J, Yi X, Zhang X, Zhong W. A chemoenzymatic strategy for site-selective functionalization of native peptides and proteins. *Science* 376;2022:1321-1327.

Site-directed modification of proteins can be very useful for changing their function. Such modification is generally approached by total chemical synthesis or genetic incorporation of unnatural amino acids. Fryszkowska *et al.* adopt an alternative approach: enzyme-catalyzed conjugation that relies upon site-selectivity of an enzyme. The authors are interested in making modified forms of insulin to improve characteristics for the treatment of diabetes mellitus. Insulin has 3 primary amino groups, any of which may be targeted for modification: the *N*-terminal α-amino groups on the A and B chains, and the ε-amino group on the side-chain of B-chain lysine-29. The reactivity of these groups is difficult to separate on the basis of *p*Ka alone. The authors first derivatize all three amines with a phenylacetyl protecting group using an N-hydroxysuccinimide-ester. They then test penicillin G acylase enzymes from various sources for their ability to

deprotect the three residues in a regioselective manner. They perform directed evolution to improve the specificity of the best candidates. A complete series of the possible combinations of protected amines is finally achieved by adding a regiospecific catalysis of the reverse reaction – transfer of a phenylacetyl group form a phenylacetyl ester to an amine. With the various combinations of protected groups as substrate, conjugation of useful functional groups may be installed at any combination of positions. After conjugation of functional groups, the protecting groups must be removed. The authors perform further rounds of directed evolution to yield enzymes for this purpose. The methodology enables the synthesis of insulin derivatives with diverse capabilities, including extended serum half-life, partial insulin receptor activation, glucose responsiveness, and regulation of insulin availability.

Cao L, Coventry B, Goreshnik I, Huang B, Sheffler W, Park J S, Jude K M, Marković I, Kadam R U, Verschueren K H G, Verstraete K, Walsh S T R, Bennett N, Phal A, Yang A, Kozodoy L, Dewitt M, Picton L, Miller L, Strauch E-M, Debouver N D, Pires A, Bera A K, Halabiya S, Hammerson B, Yang W, Bernard S, Stewart L, Wilson I A, Ruohola-Baker H, Schlessinger J, Lee S, Savvides S N, Garcia K C, Baker D. Design of protein-binding proteins from the target structure alone. ***Nature*** **605;2022:551-560.**

This paper represents a step toward the ultimate goal of designing high-affinity binders to selected sites on protein surfaces of known structure *de novo*, using computational tools alone, without the need for prior information about existing binding partners or iterative empirical optimization. The ambitiousness of this goal may be recognized when recalling that many proteins lack obvious surface pockets or clefts, and have few privileged side-chains to target. The authors' strategy initially comprises 3 steps: (1), compilation of a comprehensive list of (individually weak) binding affinities with billions of disembodied amino acids, emphasizing non-polar interactions; (2), identification of a large set of mini-protein backbones (50-65 amino acids in length) that can host many of these interactions without clashing with the target; and (3), identification of recurrent backbone motifs within this set of mini-proteins. The interface structural motifs identified in this way are then subjected to a more focused search involving another round of docking and design. Using this approach, the authors design stable binding proteins for 14 sites on 12 diverse protein targets, and demonstrate nanomolar to picomolar binding affinities following experimental optimization. Crystal structures for 5 of the binder-target complexes are solved and shown to accord with the design models. The methodology still generates many designs that don't bind, struggles to identify binders for polar target sites, and requires experimental optimization of binders. However, the advance that the present work represents encourages optimism that artificial intelligence methods will play an increasingly significant role in the design of protein binders for therapeutic purposes in the future.

MASS SPECTROMETRY

Krenkel H, Brown J, Richardson K, Hoyes E, Morris M, Cramer R. Ultrahigh-throughput sample analysis using liquid atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. ***Analytical Chemistry*** **94;2022:4141-4145.**

Radosevich A J, Pu F, Chang-Yen D, Sawicki J W, Talaty N N, Elsen N L, Williams J D, Pan J Y. Ultra-high-throughput ambient MS: direct analysis at 22 samples per second by infrared matrix-assisted laser desorption electrospray ionization mass spectrometry. *Analytical Chemistry* 94;2022:4913-4918.

Two groups demonstrate methods for ultra-high-throughput mass spectrometry. The capability to process samples very rapidly is of special interest for population diagnostic analyses, enzyme kinetic and pharmacokinetic analyes, and screening of compound libraries in the pharmaceutical industry. Previously, speeds up to 6 samples/s have been achieved with acoustic droplet ejection. Krenkel *et al.* here demonstrate speeds of 20 samples/s in peptide analysis and 40 samples/s in an enzyme assay. They employ liquid atmospheric pressure matrix-assisted laser desorption/ionization (LAP-MALDI). In this technique propylene glycol is added to a solution of the MALDI matrix compound (α-cyano-4-hydroxy-cinnamic acid) in an acetonitrile/water mixture with a final concentration of propylene glycol of 60% by volume. The liquid matrix supports high ion signal stability, produces multiply charged analyte ions similar to electrospray ionization (ESI), and provides low matrix background. It is inherently fast, with ion packets from individual desorption events <5 ms wide. Samples are supplied in a microtiter plate format. Required sample volume is small (<2 µL), and <1% is usually consumed. Radosevich *et al.* achieve 22.7 samples/s, again in microtiter plate format. They employ infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI). This is a hybrid, ambient pressure ionization technique that combines MALDI and ESI: laser desorption is followed by post-desorption ESI. The matrix required is water or ice. No additional sample preparation is therefore needed for tissues, biological fluids or biochemical reactions, and ion suppression due to salts or detergents is minimized.

PROTEOMICS

Sun X, Sun H, Han X, Chen P-C, Jiao Y, Wu Z, Zhang X, Wang Z, Niu M, Yu K, Liu D, Dey K K, Mancieri A, Fu Y, Cho J-H, Li Y, Poudel S, Branon T C, Ting A Y, Peng J. Deep single-cell-type proteome profiling of mouse brain by nonsurgical AAV-mediated proximity labeling. *Analytical Chemistry* 94;2022:5325-5334.

Sun *et al.* adapt the technique of enzyme-catalyzed proximity labeling, originally developed for investigation of subcellular co-localization of proteins, as a strategy for proteome profiling of hard-to-isolate cell types. They use TurboID, an engineered biotin ligase, to rapidly biotinylate proteins localized within a specific cell type. The gene encoding TurboID is delivered to live animals by a recombinant adeno-associated virus (AAV) and expressed under control of cell-type-specific promotors. (Selective tropism of the virus may also contribute to the specificity of labeling.) The enzyme biotinylates proteins within the cell when supplied with exogenous biotin. Labeled proteins are purified, subjected to on-bead digestion, and identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The proteins may be quantified by tandem-mass-tag (TMT) labeling. The authors deploy this procedure for profiling the proteomes of mouse neurons and astrocytes. AAVs are delivered to the brain by direct venous injection. The AAVs also encode mCherry, a fluorescent marker that the

authors use to confirm that expression is occurring in the correct cell type using co-immunostaining with appropriate cell markers. Proteins and pathways selectively enriched in neurons and astrocytes are identified. This methodology affords the major advantage of eliminating the need for isolation of the cell types being studied. In principle, it is widely applicable to the study of proteomes in diverse cell-types.

FUNCTIONAL GENOMICS/PROTEOMICS

Katrekar D, Yen J, Xiang Y, Saha A, Meluzzi D, Savva Y, Mali P. Efficient in vitro and in vivo RNA editing via recruitment of endogenous ADARs using circular guide RNAs. *Nature Biotechnology* 40;2022:938-945.

Yi Z, Qu L, Tang H, Liu Z, Liu Y, Tian F, Wang C, Zhang X, Feng Z, Yu Y, Yuan P, Yi Z, Zhao Y, Wei W. Engineered circular ADAR-recruiting RNAs increase the efficiency and fidelity of RNA editing in vitro and in vivo. *Nature Biotechnology* 40;2022:946-955.

Correction of G-to-A point mutations and premature stop codons in DNA by the editing of RNA sequence rather than DNA itself is an attractive approach to the treatment of genetic disease. Its advantage is that the intervention is not heritable, so it avoids risks inherent in changes to the germline. RNA editing may be performed by conversion of adenosine to inosine by an adenosine deaminase acting upon RNA (ADAR). ADAR acts on double-stranded RNA, so site-specificity is programmable by supplying a guide RNA of appropriate sequence. The product, inosine, is read as guanosine during subsequent splicing and translation. Various ADARs have been engineered to improve efficiency of this process, but off-target A-to-I conversion remains a problem. Two groups now demonstrate major improvements in the efficiency, durability and specificity of RNA editing by replacement of linear RNA with circular RNA to act as a guide RNA for ADAR recruitment, taking advantage of its extended half-life resulting from resistance to endogenous cellular exonucleases. Linear RNAs may be circularized by a previously published method in which the RNA is supplied with flanking sequences for twister ribozymes that undergo autocatalytic cleavage, leaving termini that are ligated by a ubiquitous, endogenous RNA ligase. Yi *et al.* control bystander, off-target editing within the target transcript itself by completely eliminating from the guide sequence uridine nucleotides opposite off-target adenosines. Katrekar *et al.* introduce 8–12-bp loops upstream and downstream of the target adenosine to create A-specific bulges for the same purpose. Katrekar *et al.* additionally find that recruitment of endogenous ADAR supports robust *in vivo* editing without needing to supply an ADAR exogenously. The improvements in RNA editing observed with the use of circularized RNA suggest that circular RNA might also find utility in RNA inhibition, antisense oligonucleotide, and CRISPR-Cas editing.

Wang J, He Z, Wang G, Zhang R, Duan J, Gao P, Lei X, Qiu H, Zhang C, Zhang Y, Yin H. Efficient targeted insertion of large DNA fragments without DNA donors. *Nature Methods* 19;2022:331-340.

Anzalone A V, Gao X D, Podracky C J, Nelson A T, Koblan L W, Raguram A, Levy J M, Mercer J a M, Liu D R. Programmable deletion, replacement, integration and inversion of large DNA sequences with

twin prime editing. ***Nature Biotechnology*** **40;2022:731-740.**

Our best hope for definitive treatment of genetic diseases is to perform specific, targeted deletion of defective DNA sequences and insertion of functional sequences large enough to replace entire coding regions of pathogenic variants. The work of two groups exemplifies new, rapid progress toward fulfillment of this aspiration. Methods requiring the introduction of double-stranded breaks aren't precise or efficient enough for the task, but advances in the technique of prime editing are showing promise. Prime editing involves DNA nicking but not double-strand breakage. The prime editing protein comprises a catalytically impaired Cas9 nickase and a wild-type (PE1) or engineered (PE2) reverse transcriptase. A prime editing guide RNA (pegRNA) that contains an RNA template as a contiguous extension of the guide RNA is supplied. Nicking by Cas9 exposes a 3′ DNA flap, which binds to the primer binding site of the RNA template, serving as a primer for the reverse transcriptase. The reverse transcriptase proceeds to extend the 3′ flap by copying the edit sequence of the pegRNA. After reverse transcription, the newly synthesized 3′ flap containing the edited sequence invades the adjacent DNA to replace the redundant 5′ flap sequence. The opposite, non-edited strand is then repaired using the edited strand as a template. Two groups now show that an augmented version of this technique is capable of integrating much longer DNA sequences than hitherto possible. They use 2 pegRNAs, to template the synthesis of complementary DNA flaps on opposing strands of genomic DNA so that the entire sequence between the pegRNA nick sites is replaced. Because both edited strands are synthesized by prime editors, the requirements for strand invasion and second-strand repair – processes that might otherwise revert the edited sequence to its original form – are bypassed. Wang *et al.* demonstrate targeted insertion of 20 bp to ∼1 kb of DNA with efficiencies that reach ∼60% for 100-bp and ∼30% for 250 bp. Anzalone *et al.* show insertion of a 108-kb sequence with an efficiency of 16%. Anzalone further achieve targeted inversion of a 40-kb sequence using the editing technique to insert recombinase recognition sites at different locations in a programmed manner. This result indicates that targeted integration of gene-sized DNA segments is practicable.

Sánchez-Rivera F J, Diaz B J, Kastenhuber E R, Schmidt H, Katti A, Kennedy M, Tem V, Ho Y-J, Leibold J, Paffenholz S V, Barriga F M, Chu K, Goswami S, Wuest A N, Simon J M, Tsanov K M, Chakravarty D, Zhang H, Leslie C S, Lowe S W, Dow L E. Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. ***Nature Biotechnology*** **40;2022:862-873.**

Kim Y, Lee S, Cho S, Park J, Chae D, Park T, Minna J D, Kim H H. High-throughput functional evaluation of human cancer-associated mutations using base editors. ***Nature Biotechnology*** **40;2022:874-884.**

In oncology there is a need to distinguish mutations that contribute to the malignant phenotype (driver mutations) from functionally neutral (passenger) mutations. Similarly, there is a need for information about the effects of variants of unknown significance upon which to base clinical management decisions for individual patients. For assistance in solving these problems, two groups now make an improvement in base editing

methodology for screening the effects of different point mutations in selected proteins at scale. Base editing, like prime editing, avoids making double-stranded breaks in DNA by using a Cas9 nickase. The nickase is fused to a deaminase that generates C-to-T or A-to-G transversions at positions directed by a guide RNA. In pooled screens with guide RNA libraries, hits are usually identified on the basis of a change in abundance of the cognate guide RNA as determined by high-throughput sequencing. However, variation between guide RNAs in editing efficiency may also contribute to such a signal. Therefore, Sánchez-Rivera *et al.* and Kim *et al.* both incorporate into their vectors a 'sensor' of the same sequence as the target site to act as a base editing surrogate. Sequencing the sensor cassette provides a measure of the efficiency of base editing at cognate sequences. Both groups find good correlation between base editing at the target and proxy sequences. The use of sensors improves the accuracy of the base editing screens.

Erwood S, Bily T M I, Lequyer J, Yan J, Gulati N, Brewer R A, Zhou L, Pelletier L, Ivakine E A, Cohn R D. Saturation variant interpretation using CRISPR prime editing. *Nature Biotechnology* 40;2022:885-895.

Erwood *et al.* also contribute to the technology available for saturation mutagenesis screening, but employ a prime editing approach instead of base editing. They address the problem of determining the effect of mutations in diploid cells in which the unaffected allele can mask the phenotype associated with induced mutations. Previous studies have often employed HAP1, a near-haploid cell line, for saturation mutagenesis screening, but these cells may revert to diploid, and in any case restrict the scope of studies. In order to be able to screen diploid cells, Erwood *et al.* instead use paired guide RNAs specific for both naturally occurring alleles at loci of interest, one guide RNA to delete a large fragment the genomic region encompassing the target allele and the other guide RNA to mutate the remaining allele. They deploy this strategy in a study of *NPC1*, the gene encoding intracellular cholesterol transporter 1, mutations in which are known to cause the lysosomal storage disorder Niemann-Pick syndrome type C. Monitoring lysosomal expansion as an index of such pathology in a pooled assay using fluorescence-activated cell sorting, they identify 410 of 706 missense mutations in *NPC1* that produce a deleterious phenotype.

MACROMOLECULAR CHARACTERIZATION

Stiller J B, Otten R, Häussinger D, Rieder P S, Theobald D L, Kern D. Structure determination of high-energy states in a dynamic protein ensemble. *Nature* 603;2022:528-535.

The three-dimensional (3D) structure of many proteins is dynamic: the polypeptide conformation makes excursions into alternative, high-energy states as part of the duty cycle by which the protein performs its function. Such conformations are generally sparsely populated, and usually cannot be trapped, so their experimental elucidation is challenging. Stiller *et al.* describe methodology for characterizing the thermodynamics, kinetics and high-resolution structure of high-energy states based on nuclear magnetic resonance (NMR) spectroscopy. They employ an NMR technique known as Carr-Purcell-Meiboom-Gill

(CPMG) relaxation dispersion, an approach to obtaining information about dynamic processes that occur in the time-frame of µs to ms, including ligand binding/release, folding/unfolding events, allostery, and catalytic turnover. The authors couple this approach to the use of paramagnetic centers that induce so-called pseudocontact shifts (PCS), in which paramagnetic metals magnify CPMG relaxation dispersion profiles to provide long-range spatial information up to 50 Å, thereby providing a capability to analyze proteins up to 60 kDa. The authors apply their methodology to the catalytic cycle of adenylate kinase. This enzyme transfers phosphate between a pair of ADP molecules to yield ATP and AMP. Their data indicate the existence of a high-energy structure primed for substrate binding or product release and a catalytic mechanism that involves transition from this state to a closed state by induced fit in which the phosphate transfer reaction occurs. The PCS-CPMG methodology is shown to be generally applicable to diverse proteins and is expected to contribute to future studies of high-energy proteins states.

IMAGING

Glaser A K, Bishop K W, Barner L A, Susaki E A, Kubota S I, Gao G, Serafin R B, Balaram P, Turschak E, Nicovich P R, Lai H, Lucas L a G, Yi Y, Nichols E K, Huang H, Reder N P, Wilson J J, Sivakumar R, Shamskhou E, Stoltzfus C R, Wei X, Hempton A K, Pende M, Murawala P, Dodt H-U, Imaizumi T, Shendure J, Beliveau B J, Gerner M Y, Xin L, Zhao H, True L D, Reid R C, Chandrashekar J, Ueda H R, Svoboda K, Liu J T C. A hybrid open-top light-sheet microscope for versatile multi-scale imaging of cleared tissues. *Nature Methods* 19;2022:613-619.

Light-sheet microscopy has emerged as the method of choice for deep-tissue imaging of cleared tissues when high resolution and contrast are needed for volumetric imaging. However, different applications and clearing methods require different microscope configurations, so study design is often limited by the instrumentation available. Glaser *et al.* describe a new instrument configuration that provides a more favorable balance of performance and versatility by fulfilling requirements for: user-friendly mounting of multiple specimens in standard holders, compatibility with diverse clearing methods, absence of limits on specimen size, large imaging depth to accommodate samples of appreciable thickness, and capability for screening large (cm-scale) volumes at moderate (µm) resolution followed by detailed study of localized regions of interest (mm-scale) at sub-µm resolution. The authors adopt an open-top configuration in which the optical components are situated underneath the microscope stage for maximum flexibility in the type of sample situated on the stage top. Three objectives are employed on the same instrument: one is used for light-sheet illumination, the second for orthogonal dual-objective (ODO), and the third for non-orthogonal dual-objective (NODO) imaging. In this hybrid optical architecture, the ODO path is used for low-magnification imaging. To suppress aberrations caused by refractive index mismatching along the light path, a solid immersion meniscus lens (SIMlens) is employed for the angled ODO objective. This covers all existing clearing protocols. The NODO path is used for high-resolution imaging. A vertically oriented multi-immersion objective with long working distance is

used for the NODO path, providing good imaging depth and high tolerance for refractive index differences. The versatility of this system will afford significant advantages in the setting of core facilities.

CELL BIOLOGY

Moore D F, Sleat D E, Lobel P. A method to estimate the distribution of proteins across multiple compartments using data from quantitative proteomics subcellular fractionation experiments. *Journal of Proteome Research* 21;2022:1371-1381.

This paper describes freely available software for calculating the distribution of proteins among subcellular compartments based upon proteomic study of subcellular fractions. Distributions of each protein are calculated from quantitative mass spectrometric data for subcellular fractions, *e.g.* using TMT reporter ions. These distributions are compared with a set of marker proteins each of which is uniquely assigned to a single compartment. The method requires accurate measurement of the total protein in the sample before subcellular fractionation, and the total protein in each of the fractions, so that the amount of each protein relative to the total protein can be calculated. For the purposes of accurate calculation of the distribution of individual proteins between compartments, the mass spectrometric data are used to calculate the enrichment or depletion of each protein in each fraction compared to the initial tissue or cell homogenate. Importantly, this methodology is suitable for quantifying the subcellular distribution of the many proteins that occur in more than one subcellular compartment. The authors additionally provide guidance on how to use peptide-level data for investigation of the distribution of different isoforms of a protein among compartments, how to assess the accuracy of compartment assignments when different fractions might contain material from more than one subcellular compartment, and how to recognize proteins that may be localized in a compartment that lacks a marker protein.

Cho N H, Cheveralls K C, Brunner A-D, Kim K, Michaelis A C, Raghavan P, Kobayashi H, Savy L, Li J Y, Canaj H, Kim J Y S, Stewart E M, Gnann C, Mccarthy F, Cabrera J P, Brunetti R M, Chhun B B, Dingle G, Hein M Y, Huang B, Mehta S B, Weissman J S, Gómez-Sjöberg R, Itzhak D N, Royer L A, Mann M, Leonetti M D. OpenCell: endogenous tagging for the cartography of human cellular organization. *Science* 375;2022:eabi6983.

Eukaryotic protein interactions have historically been investigated using ectopic expression combined with the yeast two-hybrid system, or epitope tagging combined with immunoprecipitation and mass spectrometric protein identification. Protein subcellular localization, an expression of broader scale protein interactions, has been investigated using fluorescence imaging. In a seminal study, Cho *et al.* combine investigation of protein interaction and protein localization, gaining unexpected insights into cellular organization. They employ CRISPR gene editing to introduce fluorescent tags into proteins of human embryonic kidney (HEK) 293T cells. For tagging they use the split-mNeonGreen$_2$ system, in which the fluorescent protein is separated into 2 fragments: a short mNG11 fragment that is added to the target protein by genetic engineering and the larger,

$mNG_2$1-10 portion that is expressed in *trans*. The 2 fragments complement one another to create the completed $mNeonGreen_2$ fluorescent tag. A library of cell lines expressing a total of 1310 tagged proteins individually in concentrations detectable by fluorescence imaging is assembled. Importantly, the genes remain under native expression regulation. The fluorescent tag also serves as 'bait' for immunoaffinity isolation of protein complexes for mass spectrometric identification of interacting proteins following solubilization of complexes with digitonin detergent in the presence of benzonase, a nuclease for DNA and RNA. In total, 29,922 interactions are identified among 5292 proteins ('baits' and 'preys'). The authors delineate stable interactions by estimating the stoichiometry of bait and prey proteins in immunoprecipitates: a high abundance of the prey protein relative to the bait protein is taken to indicate a stable, 'core' interaction. Unsupervised Markov clustering of interactions weighted according to stoichiometry reveals 'communities' of proteins whose members are presumed to be involved in functionally related pathways. Three hundred such communities are identified, comprised of a total of 2096 proteins (baits and preys). Ontology analysis confirms functional linkage. These data are used successfully to suggest functions for poorly annotated proteins. Protein subcellular localization is assessed in 3D by spinning disk confocal fluorescence microscopy. The authors manually assign each protein to subcellular compartments (55% of the proteins are detected in more than 1 compartment), but follow this up with image analysis using artificial intelligence methodology. The latter approach produces automated, self-supervised, multidimensional classification of protein distributions that is not influenced by *a priori* assumptions or manually assigned labels. The resulting distribution map identifies a rich diversity of protein 'territories' that mirror the manual annotations. Combining the results for protein interaction communities and protein distribution fine-structure then reveals not only the expected concordance between the two but the additional insight that detailed patterns of 'localization similarity' are strongly predictive of protein community membership. This suggests, rather surprisingly, that protein-protein interactions may be identified by quantitative comparison of spatial distribution alone. The authors combine their spatial- and interaction-based groupings into a single, hierarchical description of proteome organization. At the highest level of the organizational hierarchy, they distinguish 3 branches: soluble proteins, membrane proteins, and, surprisingly, a third branch consisting of RNA-binding proteins. Recalling that RNA has been removed prior to protein interaction analysis, this third branch must reflect features of protein interaction beyond mutual affinity for RNA. The authors note enrichment of disordered proteins within this branch, including proteins involved in biomolecular condensation, suggesting that condensation might play a driving role in shaping the proteome of RNA-binding proteins. Furthermore, the results overall are consistent with the notion that disorder plays as important a role in shaping proteome organization as does hydrophobicity in determining the function of membrane proteins.